Pose and Motion Estimation of a Moving Rigid Body with Few Features

Valentin Borsu and Pierre Payeur
School of Information Technology and Engineering
University of Ottawa
Ottawa, ON, Canada
[vbors100, ppayeur]@site.uOttawa.ca

***Abstract*— This paper proposes a reliable solution to the problem of estimating the motion of a rigid object moving freely in 3D space, through the use of a passive vision system. The feature-based tracking technique builds upon the selection of a consistent set of features and their tracking on a frame-by-frame basis. A thorough investigation is conducted to determine a proper vision system setup, which results in a configuration that ensures the coverage of the complete patterns of motion that the object may exhibit. While the system relies on low resolution cameras, the proposed algorithm provides subpixel accuracy on the pose estimation of the rigid body and its associated motion. The algorithm is experimentally validated and operates within an execution timeframe that makes it suitable for real-time processing applications.**

Keywords*— *feature extraction; sparse optical flow; feature correspondence; tracking; 3D reconstruction; pose estimation; motion estimation.

I. Introduction

The goal of this work is to estimate the pose and motion of a rigid object translating and rotating freely in 3D space. The motion is assumed to be smooth and continuous, while the color and texture properties of the object are not strongly contrasting or easily detectable. The proposed solution aims at providing accurate estimation of the location of the rigid body in the 3D world and its associated speed, without referring to an exact 3D CAD model of the object.

The suggested pose and motion estimation technology is meant to become an integrated part of an autonomous robotic system to work in interaction with moving automotive parts, which are rotating and translating on an assembly-line. The task is an important component of an industrial setting for automatic quality control. Therefore, the robot is considered as a central element for an on-line defects marking system. The pose and motion estimator complements other sensors that provide extremely accurate information about the surface shape characteristics of body parts. For the robot to perform actions on a part that is moving with the assembly line, the motion of the gripper must be planned in accordance with the motion of the object, therefore the requirement for accurate and real-time pose and motion estimation. By ensuring synchronization between the robot end-effector and the automotive body part, the marking operation of surface defects can be accomplished as if in a pseudo-static environment.

The current techniques used for tracking unidentified objects require the selection of some visually significant keypoints that can be easily extracted and followed over several frames. The main problem resides in how these feature points can be selected, such that they are unique or nearly-unique, and can be compared in a distinctive way to other points in subsequent images. By analyzing how the position of these features changes from frame to frame, useful data for estimating the motion, as well as a sparse structure model of the rigid body, can be obtained.

The design of a solution must take into account the specificities of the industrial setup. The starting premise was linked to the fact that merging the information from several vision sensors could bring reliable data about the complete motion of the rigid body, in any direction. However, solving the feature correspondence problem under mixed effects of large baseline, low amount of overlap, complex environment and low resolution, preempted the implementation of such a generic solution. A careful analysis on the complete set of motion patterns that can be performed by the object in the industrial perspective of an assembly line led to the selection of a setup consisting of a small baseline stereo-vision system appropriately positioned in the scene.

Finally, with the addition of extra knowledge about the general appearance of objects to be inspected on the assembly line, but without requesting an exact CAD model, a reliable algorithm for pose and motion estimation is developed. The system was tested with offline data and proved to be suitable for real-time processing.

The paper is organized as follows: Section II proposes a review of promising existing techniques to address the several challenges involved in the considered application. In Section III, two experimental multiple-view vision systems are introduced and characterized. In Section IV an original pose and motion estimation algorithm is detailed while Section V introduces the experimental validation. Finally, Section VI concludes the work and discusses potential extensions.

II. State-of-the-Art

In this section important research work on three major aspects that provide partial solution to the motion estimation problem are reviewed. The extraction of reliable features, the

978-1-4244-4778-7/09/$25.00 ©2009 IEEE

computation of optical flow, and the formal estimation of motion are examined.

A. Feature Extraction

The selection of features to track is an important aspect that must be addressed at an early stage as it directly conditions the accuracy of the motion estimates. Harris and Stephens [1] introduced the combined corner and edge detector based on the local auto-correlation function to deal with image regions that contain texture and isolated features. Inspired by the proposed approach and the specified measure of the corner quality, Shi and Tomasi [2] set up a different validation gate starting from the same data which is the auto-correlation matrix. In a thorough investigation on feature matching, Vincent and Laganière [3] used corner extraction and introduced the interesting property of "repeatability". Another effective direction in the field of feature extraction was initiated by Lowe [4] with the development of SIFT features, in which the process of extracting keypoints becomes invariant to scaling and rotation.

However, under the constraints of the industrial application considered, low-resolution cameras are privileged and therefore images do not contain a large amount of fine details. Moreover, the inspected objects present very few sharp and unique features from a visual perspective. In addition to this, both pose and motion must be estimated in real-time. Therefore, solving correspondences under these conditions with the SIFT features approach does not represent a realistic alternative.

B. Sparse Optical Flow Computation

Provided that a reliable set of features is available, velocity or displacement estimates can be extracted. For that matter, the Lucas-Kanade (LK) feature tracker [5] can be applied to the subset of points in the input image. The suggestion of Lucas and Kanade to use a "coarse-fine strategy" as a registration technique represents the starting point for the pyramidal implementation of the LK tracker, detailed in [6]. In the latter, the problem of a natural trade-off between local accuracy and robustness is also addressed, when selecting the size of the search window in the motion tracking process.

A sparse perspective over the Block Matching Algorithm is introduced by Chen *et al*. [7]. Sui *et al.* [8] present another approach for feature tracking in image sequences. Under this framework, using the projective invariant of Barrett and Payton [9] and the tracking of 8 general points in space, all other feature points that exist in the image can be tracked by means of a Hough transform technique. The low number of features required is appealing. However, under the constraints of an industrial setup, the suitability of the latest two solutions is questionable, since the trade-off between accuracy and computational costs remains critical.

C. Motion Estimation and Robotic Interaction

Huang and Netravali [10] present a review on 3D-to-3D/2D-to-2D correspondences to estimate the 3D motion and structure of rigid objects from corresponding features at different times. Weng *et al.* [11] go along the same direction and introduce an algorithm for estimating the motion parameters and structure of the scene from point correspondences between two perspective views. Conversely, the proposed error correction algorithm is computationally expensive. It also imposes constraints on the necessity of a large field-of-view, and a simplified case of motion that should approximate a translation orthogonal to the image plane, similar to a far focus of expansion.

As Holt and Netravali [12] observe, more useful information and improvement for the motion estimation and understanding can be obtained by analyzing an image sequence with more than just two frames. In their approach, they consider a motion model containing only a few parameters which can be presumed to remain constant over a short period of time. However, their formulation imposes constraints on the freedom of motion, since they consider that the translation does not vary between successive pairs of time intervals. Therefore, more insight about the motion can be obtained, but at the cost of restricting the freedom of motion of the rigid body. Kuang and Liu [13] introduce a pose estimation algorithm building upon a setup with a stereo-vision system and a rigid body equipped with a couple of known predefined markers on its surface. Their solution is based on the fact that the increments from the set of points on which rotation has been applied are perpendicular to the direction of rotation.

Overall, the literature remains very limited about the problem of robotic interaction with moving rigid bodies in an industrial setting. Yoon *et al.* [14] highlight one of the important problems faced in industrial quality control, which is the lack of a considerable amount of features to track in order to determine the motion. In their work, the authors try to estimate the complete pose of an industrial object by using its circular-shape features. Their algorithm achieves good accuracy by tracking only three of the geometrical features of the object with a camera-in-hand solution. Under the same framework, Chang *et al.* [15] propose a multiple-level robot control solution for a peg-and-hole experiment on the same moving body as in [14]. They point-out the necessity of highly accurate ground truth data for validating the pose and motion estimates of the object throughout its motion sequence. For this purpose and for calibration and visual servoing validation, they make use of a six degree-of-freedom NIST laser tracker.

The problem in the current research and the one introduced in [14] are similar from the perspective of estimating the pose without any *a priori* knowledge about the 3D CAD model of the object. Nevertheless, in order to be able to solve this problem in the case of a rigid body exhibiting only a small number of features, extra knowledge about the object structure is required. Since using an exact CAD model of the object is not desirable, only generic geometrical characteristics about the automotive part are provided to the pose and motion estimator. In that, the amount of extra information provided to the solution proposed here remains smaller than what is considered in [14]. In the latter a specific reference frame is assigned to the industrial object, based on the locations of the three circular shapes of interest, measured with the laser tracker.

III. EXPERIMENTAL STEREOSCOPIC SETUP

In order to reproduce the automotive body inspection scenario in the laboratory, an experimental setup has been built that uses a rigid body structure consisting of a mobile robot to generate the movement, along with a mock-up door mounted on its upper face, as shown in Fig. 1. The prototype was built with the objective of getting as close as possible to the actual set of visible features offered by a real automotive body part, in this case, a car door with a window, the door knob and a key hole.

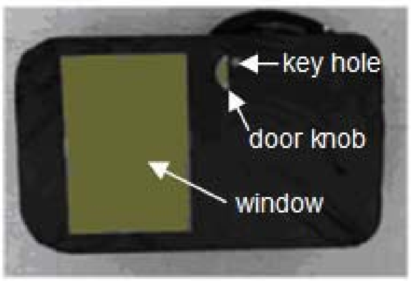

Figure 1. Mock-up rigid body used for experimental evaluation.

In the process of selecting the most suitable configuration for a multiple-view vision system in the context of on-line quality control, a number of settings have been evaluated. After a thorough investigation regarding the patterns of motion that the rigid body might exhibit on the assembly line, a configuration consisting of two orthogonal cameras was evaluated, as shown in Fig. 2a. One camera is positioned over the inspected part and points downward, while the second camera collects a lateral view of the moving part. Therefore, only two cameras are sufficient to extract the necessary motion information. As imposed by stereoscopic vision, the correspondence problem must be resolved in order to reconstruct the 3D position of some keypoints. Unfortunately, in this case the correspondence problem cannot be solved accurately. This is mainly because of the small number of features that can be extracted and matched from the limited overlapping regions of the views.

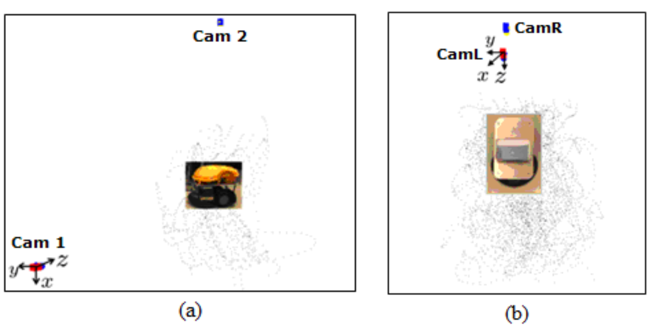

Figure 2. Alternative configurations with: a) two orthogonal cameras (Cam 1 and Cam 2), and b) a single stereo-vision system (CamR and CamL).

Considering the relatively straight motion of the assembly line, the acquisition system was modified to measure the motion information with a single stereo-vision configuration located above the assembly line and pointing downward, as shown in Fig. 2b. In this configuration the world reference frame is attached to the left camera (CamL) with the optical axis, *Z*, pointing down. The *XY* plane defines the surface of motion of the object. The motion pattern of the rigid body is mainly a translation along the *X* axis in the first part of the sequence, coupled with a minor translation along the *Y* axis in the second part of the motion sequence. The latter corresponds to a slight turn on the right w.r.t. the direction of motion. The speed of the mobile robot varies only slightly throughout the motion. Also, as the robot is manually driven by a remote control, it exhibits a low level of vibrations throughout the sequence of motion. Following experimentation, a baseline, *b*, of about 25cm between the two cameras was found to provide better results than the cases with closer vision sensors under the same configuration. The experimentation conducted on the mock-up model of a car door revealed that the stereo-vision configuration illustrated in Fig. 2b, with its principal axis pointing perpendicularly to the object and direction of motion, represents the most suitable acquisition strategy for estimating the pose and motion of the automotive body part moving on an assembly line. The calibrated vision system includes two Point Grey Flea2 IEEE-1394b CCD cameras with CCTV 3.5 mm lenses. Software solutions are implemented in C++ with the open-source library for computer vision, OpenCV.

Also, in order to create a set of ground truth data that will be compared to the pose and motion estimates, the trajectory of the mobile part is recorded during motion by tracing a line over a 156x140 cm grid (made of 1x1 cm squares) that was fixed on the floor. A third camera, synchronized with the calibrated vision system, provides information about the timing of the movement by detecting the position of the pencil marker over the grid during the recorded sequence. Although, not as highly accurate as the ground-truth data accumulation presented in [15], the recorded trajectory is fairly accurate and provides a good basis of comparison.

The data is recorded at a frame rate of 15fps, and 640x480 pixels of resolution. During this initial phase of experimentation the videos are recorded offline, but the application runs fast enough to operate in real-time on live video feeds. Since the rigid body is moving with slightly variable speed and because the frame-extraction rate is also imposed by the processing of the ground truth data, the frame-extraction process is triggered every time the pencil intersects with grid lines separated by 2cm. As a result, the average frame extraction frequency, f_{extr}, is 1.5Hz. A total number of N=48 extracted frames are employed in the sequence. The total displacement of the mock-up door is about 100 cm throughout the entire sequence. The duration of the entire processed video sequence is about 40 s.

IV. PROPOSED ALGORITHM

Building upon the last acquisition configuration detailed in Section III, an algorithm is proposed to resolve the pose and motion estimation problem. The technique offers reliable results with respect to the estimation of the pose of the object as well as for the motion it exhibits. A set of six features that can be extracted with a maximum of stability from the mock-up car door are tracked, that is the four corners of the door window and the two corners of the door knob. These keypoints are called the macro-features.

Fig. 3 presents the block-diagram of the proposed algorithm that is applied on the processed set of videos recorded with the configuration defined in Fig. 2b. Using a graphical user interface (GUI), the process is initiated by the operator who roughly selects the macro-features in the first frame that represents a full view of the inspected object in both cameras. This action (A1) is performed only once, on the frame grabbed by the right camera (camR), as an initialization step. For every chosen feature an 11x11 pixels window is defined and centered on the selected feature. A search for the exact locations of the corners (A2) is then performed only over those windows of interest to refine the estimation of the feature localization in the image. This is achieved using the Shi and Tomasi corner detector [2] with a subpixel accuracy refinement. The accurately located macro-features provide the input data to the pyramidal Lucas-Kanade tracker [6] that gives reliable information about their displacement throughout the processed sequence. Subsequently, the tracker contains a built-in validation gate for pruning away the outliers. The key elements of this procedure are the *tracking error*, linked to how much the neighborhood that contains each macro-feature changed during the tracking, and the *norm* of the optical flow in both directions with respect to the dominant norm value. Fig. 4 illustrates the sparse optical flow results for a frame captured by the right camera.

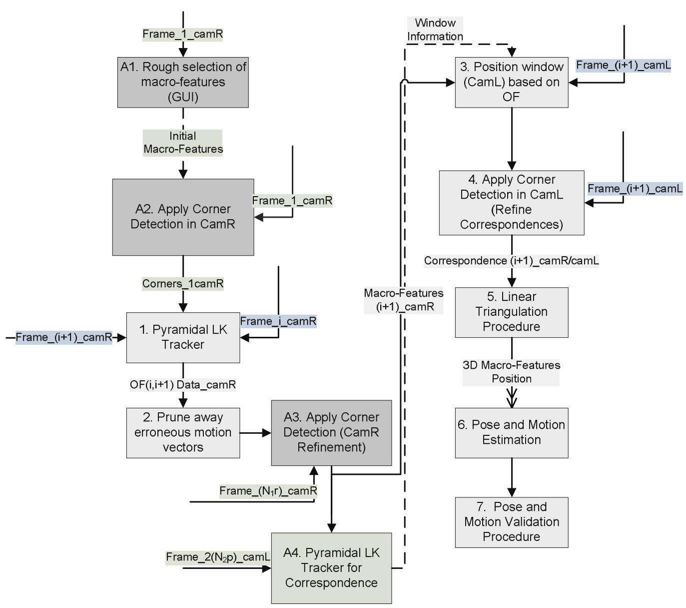

Figure 3. Proposed pose and motion estimation algorithm.

The corner detection refinement process (A3) is reapplied on every *r* frames among the set of extracted frames (i.e. N_1 times during the entire sequence, with N_1=N/r). This aims at correcting for the error that might be accumulating between the position of the features given by the tracker and the more accurate location provided by the corner detector.

Although the LK tracker is able to compute the displacements for all extracted corners, features that are not part of the selected macro-features set are discarded since they exhibit much larger variations in their position estimation over the tracking period. For that reason, for solving the correspondence problem, only the matches between the macro-features are considered.

Figure 4. Sparse optical flow data and macro-features on the rigid body.

In order to avoid computationally expensive feature matching techniques, as proposed in [3], advantage is also taken of the short baseline between the cameras. The pyramidal LK tracker also serves for directing the correspondence (A4). Also, the parameter regarding the number of levels that the image pyramid should contain can also be modified, making the proposed algorithm suitable for stereoscopic configurations with even larger baselines or systems in which the cameras are slightly tilted with respect to the principal axis of the stereo-vision sensor. The feature tracker in A4 is employed only once during the processing of the motion sequence. But in general, it can be applied every *p* frames (i.e. N_2 times during the entire processed sequence, with N_2=N/p). The feature tracker provides an initial guess of the macro-features localization in the correspondent frames.

Following a procedure similar to A1, a 12x12 pixels window is centered over each of the macro-features returned by the pyramidal LK tracker employed for the correspondence problem. As the tracker is used only once, in the next cycles of the algorithm, the positions of the windows are updated, based on the optical flow information obtained from the tracking of features in the correspondent frames taken with the right camera.

The location of the macro-features given by the pyramidal LK tracker for correspondence is being refined with the application of the Shi and Tomasi corner extractor in the corresponding frame of the left camera. The above mentioned windows, which are being displaced every cycle, represent the search regions for the corner detector. Based on the accurate estimates for the macro-features correspondences obtained previously, the 3D position of these keypoints is estimated with a linear triangulation procedure [16]. The 3D points represent the set of estimated data, to be compared with the ground truth data obtained with the help of the validation system described in Section III.

V. EXPERIMENTAL VALIDATION

Fig. 5 shows a back-projection of the 3D points over the corresponding image plane of the left camera and is used to evaluate the accuracy of the pose estimation. As a calibrated stereo-vision system is used, the epipolar lines can be computed in the frames of the left camera and used to visually inspect the location of the computed correspondences with respect to them. Experimentation demonstrated that sub-pixel accuracy is achieved on the back-projection of the estimated pose for the set containing the first four macro-features (see Fig. 5) in both directions. The last two macro-features exert subpixel accuracy on the back-projection error in the horizontal direction, and an average value of 1.5 pixels for the back-projection error in the vertical direction. As mentioned in Section III, the rigid body exhibits relatively straight and smooth motion (mainly, a translation along the *X* axis in the first part, coupled with a minor translation along the *Y* axis in the second part, w.r.t. the world reference frame as shown in Fig. 2b).

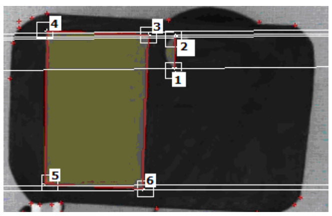

Figure 5. Image from the left camera with back-projection of estimated location of the six macro-features.

The analysis of the reconstructed 3D data is based on a set of two quality tests. The first quality measure is related to the comparison of our results with the knowledge available about the geometrical structure of the mock-up door. The reference data is accurately measured in the real world and compared to the corresponding computed values based on the 3D positions of the macro-features. The set of data consists of: the width between macro-features 3 and 4 ($width_1$), as defined in Fig. 5, between features 5 and 6 ($width_2$), and the height between macro-features 4 and 5 ($height_1$), and between macro-features 3 and 6 ($height_2$) of the rectangle representing the window. As well, the lateral size of the door knob (between macro-features 1 and 2) is monitored. The absolute relative error between the real and the estimated data is computed over the entire processed sequence. The values reported in Table I represent the mean of the absolute relative errors computed over the sequence, in the cases where corners are re-extracted every r=5 frames or not, as explained in Section IV. It can be observed that the data, in which the macro-features 5 and 6 are involved, is more erroneous than the measurements that include the other macro-features. This is caused by the higher back-projection error in the vertical direction associated with the last two-features. Moreover, from the perspective of the group containing the first four macro-features, the mean absolute error is inferior to 1cm, thus acceptable accuracy is achieved for a robotic marking process of deformed areas over an automotive part. Interestingly, there is no consistent improvement observed in the data when the features are monitored with corner re-extraction; only the rectangle width and the size of knob results tend to get closer to their real values.

TABLE I. QUALITY MEASURE 1 – GROUND TRUTH DATA COMPARISON

Corners re-extraction	**Ground Truth Comparison**				
	Width₁ Error (cm)	***Width₂ Error (cm)***	***Height₁ Error (cm)***	***Height₂ Error (cm)***	***Size of knob Error (cm)***
No	0.82	0.93	1.55	1.7	0.79
Yes	0.76	0.9	1.63	1.73	0.42

The second quality measure is related to the comparison of the pose and motion estimates with the ground-truth data recorded by tracing the trajectory of the rigid body on the grid fixed on the floor. In order to be able to compare the two different data sets, one that is measured with respect to the world reference frame (CamL) and the other one with respect to the grid, a two-step transformation procedure is employed. First, ten corner-shaped markers placed over the grid are recovered with the corner extractor. The method of Arun *et al.* [17] is used to compute the transformation matrix between the world reference frame and the grid reference frame. With the use of the estimated transformation matrix, the 3D locations of the macro-features can be expressed with respect to the reference frame attached to the grid. In the second step, the displacements in *X*, *Y* and Z directions (w.r.t. the grid reference frame) between the macro-features and the tip of the pencil tracing the trajectory are estimated manually to allow a comparison of trajectories in the grid reference frame.

The validation procedure includes two comparisons for validating the accuracy of the pose and motion estimation. Fig. 6 illustrates the real and the estimated 3D trajectories of the rigid body throughout the sequence. For computing the estimated trajectory, we have applied the transformation procedure to the first macro-feature since it exhibits the smallest back-projection error throughout the sequence. For the initial part of the processed video, where the rigid body is smoothly translating along the *X* axis, the two trajectories are very close to each-other. In the second part (for *X*>40cm), a slight divergence starts to appear between them.

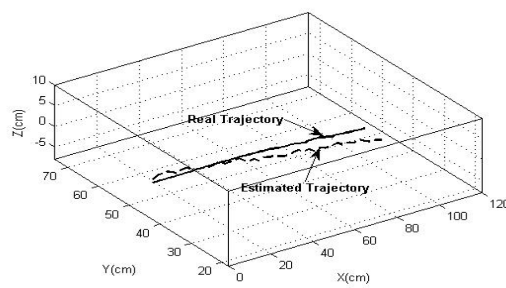

Figure 6. Real and estimated 3D trajectories comparison for the rigid body througout the motion.

The maximum value of this divergence is 3.5cm along the *X* axis, 2.8cm along the *Y* axis, and 2cm with respect to the *Z* axis. One of the most important causes of this error is the fact that the pencil that marks the grid is mounted on a spring which tends to slightly bias the trajectory when changing direction. The least-squares method used in computing the transformation matrix between the world reference frame and the grid reference frame, together with the manual measurements made in the second step of the transformation procedure, also introduce a small amount of error in the system.

For the second comparison the velocities along the *X* and *Y* axis are compared, between the ground-truth and the estimated data. In Fig. 7a it can be observed that the real and estimated velocities along the *X* axis are very close to each other. The rudimentary trajectory marking system is the cause of the spikes present in the last part of the sequence in both Fig. 7a and 7b. Moreover, the velocity in the *X* direction is not constant as the part was manually moved with a mobile robot during experimentation, which resulted in slight variations around a mean value of 3cm/s. The velocity along the *Y* axis is very small in the first part of the trajectory and is basically caused by the merged effects of noise, imperfect orthogonality of the stereo-cameras with the horizontal surface over which the movement occurs, and the small variations during the motion. Then, in the second part of the motion sequence, the absolute value of the velocity along the *Y* axis increases as the rigid body slightly turns right from the direction of motion.

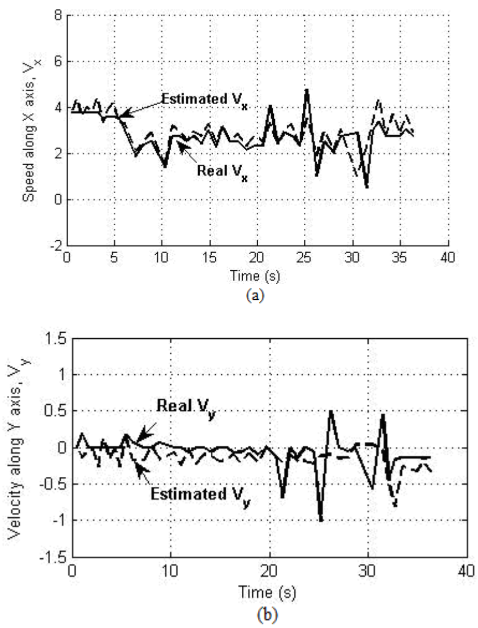

Figure 7. Real and estimated velocities (cm/s) along: a) the X axis; b) the Y axis, throughout the motion cycle.

Finally, the time performance of the entire processing per pair of frames represents approximately $\delta_p \approx 0.2$s which, subtracted from the average frame-extraction rate $\delta t \approx 0.667$s, leaves a reserve of $\delta_r \approx 0.47$s until the next frame will be captured. Experimentation demonstrated that increasing the frame-extraction rate does not significantly improve the accuracy for this type of trajectory. This makes the proposed solution a valid approach for a real-time implementation, as required in an industrial setting.

VI. Conclusions and Future Work

In this paper, the problem of pose and motion estimation of a rigid body exhibiting a limited number of dominant features, and moving in 3D space is addressed. The use of a calibrated stereo-vision system is evaluated in regards of industrial inspection systems requirements. An algorithm is proposed to effectively track the few features available and estimate pose and motion of the moving rigid body in real-time. The technique is validated experimentally and demonstrates that reliable results can be achieved, even when standard video cameras are used.

Future investigation will aim at further improving the accuracy of the pose and motion estimates, by taking advantage of the remaining time available within typical acquisition frame rates. The integration of the pose and motion estimator with a robotic arm will also be performed to provide actual interaction with the moving object under visual guidance.

Acknowledgment

The authors acknowledge the financial support from Precarn Inc., and the collaboration of Neptec Design Group Ltd and Honda Canada to this research.

References

[1] C. Harris and M. Stephens, “A combined corner and edge detector,” *Proc. of the 4th Alvey Vision Conference*, UK, 1988, pp. 147–151.

[2] J. Shi and C. Tomasi, “Good features to track,” *Proc. of 9th IEEE Conf. on Computer Vision and Pattern Recognition*, USA, 1994, pp. 593-600.

[3] E. Vincent and R. Laganière, “Matching feature points in stereo pairs: a comparative study of some matching strategies,” in *Machine Graphics & Vision*, vol. X, no. 3, 2001, pp. 237–259.

[4] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *Intl. Journal of Computer Vision*, vol. 60, no. 2, 2004, pp.91-110.

[5] B. D. Lucas and T. Kanade, “An iterative image registration technique with an application to stereo vision,” *Proceedings of the 1981 DARPA Imaging Understanding Workshop*, 1981, pp. 121-130.

[6] J. –Y. Bouguet, “Pyramidal implementation of the Lucas Kanade feature tracker – description of the algorithm,” [on-line], http://robots.stanford. edu/cs223b04/algo_tracking.pdf.

[7] Y. K. Chen, Y. T. Lin, and S. Kung, “A feature tracking algorithm using neighborhood relaxation and multi-candidate pre-screening,” *Proc. of the Intl. Conf. on Image Proc.*, vol. II, Switzerland, 1996, pp.513-516.

[8] H. Sui, Z. Zhang, and S. Kong,“Feature tracking from an image sequence using geometric invariants,” *Proc. of the IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition*, Puerto Rico, 1997, pp. 244-249.

[9] E. Barrett and P. Payton, “General Methods for Determining Projective Invariants in Imagery”, CVGIP:Image Underst*anding*, vol. 53, 1991, pp. 46-65.

[10] T. Huang and A. Netravali, “Motion and structure from feature correspondences:A Review”, *IEEE Proc.*, vol. 82(2), 1994, pp. 252-267.

[11] J.Weng, T.Huang, and N.Ahuja,“Motion and structure from two perspective views:Algorithms, error analysis and error estimation” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. II(5),1989, pp. 451-476.

[12] R. Holt and A. Netravali, “Number of solutions for motion and structure from multiple frame correspondence,” *International Journal of Computer Vision*, vol. 23(1), 1997, Kluwer Acad. Publishers, pp. 5-15.

[13] J. M. Kuang and M. Liu, “A novel explicit pose estimation algorithm based on Euclidean geometry,” *Proceedings of the 2007 Intl Conference on Information Acquisition*, 2007, Korea, pp. 247-252.

[14] Y. Yoon, G. DeSouza, and A. Kak, “Real-time tracking and pose estimation for industrial objects using geometric features,” *Proc. of the IEEE Intl. Conf. on Robotics & Automation*, Taiwan, 2003, pp. 3473-3478.

[15] T. Chang, T. Hong, G. Holguinm J. Park, and R. Eastman, “Dynamic 6DOF metrology for evaluating a visual servoing system”, *Proc. of the 2008 Performance Metrics for Intelligent Systems (PerMIS) Workshop,* 2008, USA, pp.173-180.

[16] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK, 2000.

[17] K.S. Arun, T.S. Huang, and S.D. Blostein, “Least-Squares fitting of two 3-D point set”, *IEEE Transactions Pattern Analysis and Machine Intelligence*,vol.9,no.5, 1987, pp. 698-700.